AI-Powered Offline Proctoring for Paper-Based Examinations: A Deep Learning Approach

Abstract : Academic integrity in conventional paper-based tests poses a long-standing challenge for educational institutions. Conventional monitoring methods, such as manual vigilance, are vulnerable to human weaknesses, leading to instances of academic dishonesty that go undetected. In addition, online proctoring technologies in real time are not practicable in offline test settings.We present a comprehensive comparative analysis of multiple state-of-the-art models, including **LSTM+CNN** (for temporal anomaly detection), **YOLOv8** (baseline object detection), **YOLOv8+ECA** attention (enhanced channel-wise focus), **YOLOv8+mixed attention** (hybrid spatial-channel attention), and **Vision Transformers (ViT)** (global context modeling). Experimental results demonstrate superior performance of attention-enhanced models, with YOLOv8+ECA.nd YOLOv8+mixed attention attaining **88% and 87.6%** accuracy, outperforming the baseline YOLOv8 (64%) and ViT (81.4%).

Introduction: The prevalence of academic dishonesty in paper-based examinations exposes the shortcomings of traditional proctoring methods. Educational institutions face mounting challenges as manual vigilance often fails to detect subtle or coordinated cheating behaviors, especially in large exam halls. This study introduces an AI-powered offline proctoring system that leverages advanced deep learning models—including LSTM+CNN, YOLOv8, and Vision Transformers—to autonomously monitor and detect cheating without the need for internet connectivity. By utilizing a diverse, annotated dataset of real exam scenarios, our approach surpasses conventional proctoring in accuracy, speed, and scalability. The system streamlines exam monitoring, empowering institutions to uphold academic integrity and adapt to evolving challenges in assessment environments.

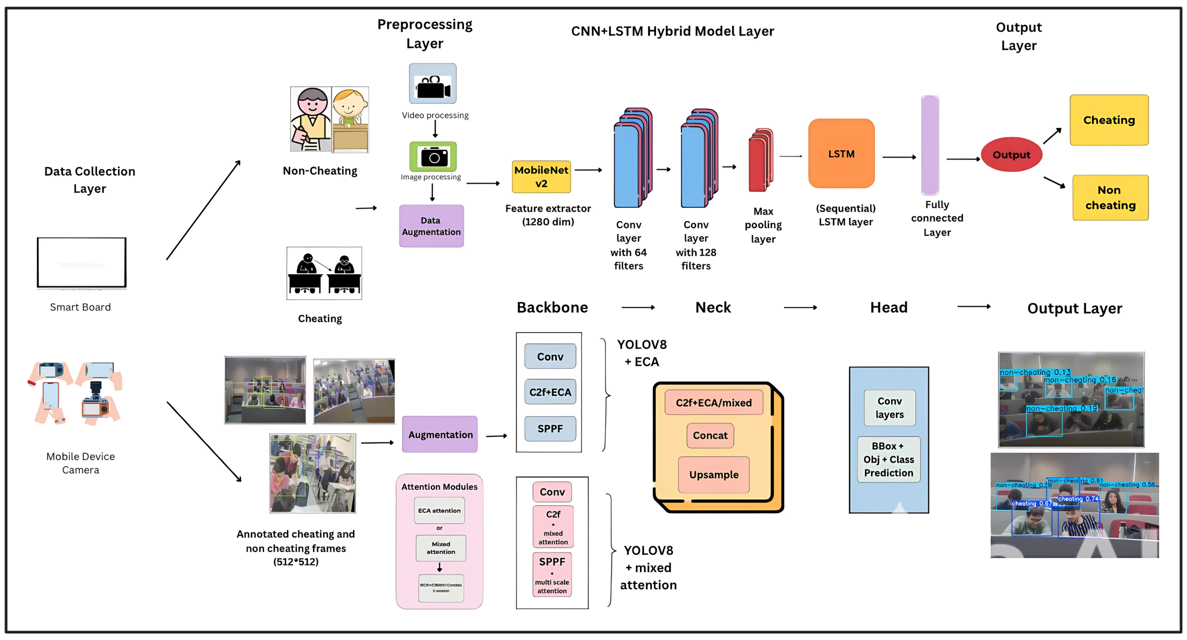

Figure 01 : Architecture Diagram

Figure 02 : Various cheating and non cheating scenarios

Methodology: This study proposes an AI-driven offline cheating detection system using deep learning models.

For the **LSTM+CNN** approach, a custom dataset of 40 videos (20 cheating, 20 non-cheating) was used. Data augmentation techniques such as rotation, zoom, brightness adjustment, and horizontal flipping were applied to improve model generalization..

The **YOLOv8** model used 3,522 manually annotated frames (cheating/non-cheating) via Roboflow, split into 70% for training, 20% for validation, and 10% for testing. Augmentation methods, including color jittering, geometric transformations, and flipping, were used. Variants of YOLOv8 were implemented with normal, **Efficient Channel Attention (ECA)**, and mixed attention mechanisms (including **CBAM**, **multi-scale attention**, and a **context-aware module**) to enhance detection performance.

For the **Vision Transformer (ViT)** model, the dataset was converted to **COCO JSON format** for compatibility, and performance was evaluated using accuracy, precision, recall, and F1-score. **YOLOv8's performance** was measured using **mAP@50** and **mAP@50-95** to evaluate detection accuracy across different IoU thresholds.

Metric	LSTM+CNN	YOLOv8 Normal	YOLOv8+ECA	YOLOv8+Mixed	ViT
Accuracy (%)	96.67	-	-	-	81.42
Precision	0.98	0.61 (overall)	0.837 (cheat)	0.826 (overall)	0.875
Recall	0.96	0.63 (overall)	0.826 (cheat)	0.792 (overall)	0.814
F1-Score	0.97	-	-	-	0.827
mAP@50	-	0.64	0.888	0.876	-
mAP@50-95	-	0.34	0.606	0.545	-
Training Time	20 epochs	Full 300 epochs	288 epochs	300 epochs	15 epochs
Inference Speed	-	4.2ms/img	4.2ms/img	4.2ms/img	-
Class Balance	36:54	1194:1802	1194:1802	1194:1802	248:75

Figure 03 : Performance metrics of various models

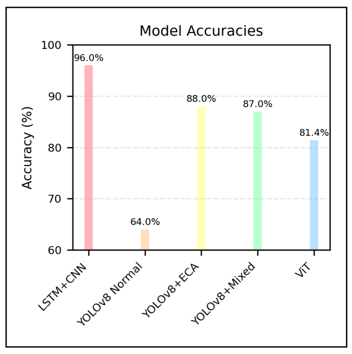

Figure 04 : Comparison of accuracies

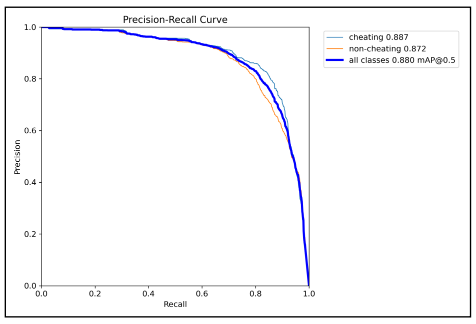

Figure 05 : Precision Recall curve of YOLOV8+ECA

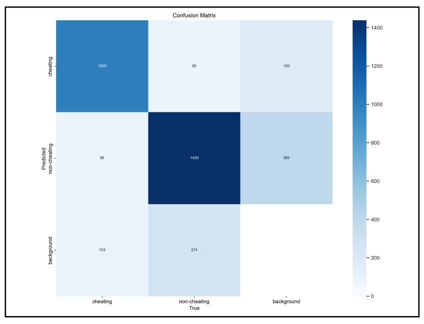

Figure 06 : Confusion matrix of YOLOV8+ECA

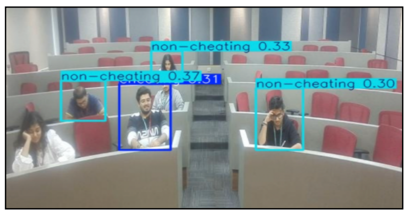

Figure 07 : Final predictions with bounding box

Comparative Analysis : The evaluated models showcased distinct strengths suited to different proctoring scenarios. LSTM+CNN excelled in accuracy and temporal behavior analysis but lacked real-time efficiency. YOLOv8+ECA emerged as the most balanced model, offering 88% accuracy with fast inference, making it ideal for real-time cheating detection. The Efficient Channel Attention (ECA) mechanism refined feature selection by focusing on the most relevant feature channels, enhancing sensitivity to subtle cheating cues while minimizing computational overhead. YOLOv8+Mixed provided strong generalization with stable performance by integrating CBAM and context-aware modules, which improved spatial-temporal attention and scene understanding. The baseline YOLOv8 lagged in precision, while Vision Transformers (ViT) contributed contextual understanding but were limited by slower training and class imbalance issues. Overall, YOLOv8+ECA stood out as the most practical choice for real world edge deployable proctoring systems.

Results : The proposed AI-based proctoring system outperformed manual and rule-based approaches, with YOLOv8+ECA achieving the best performance—mAP@50 of 0.888, precision of 0.837, and recall of 0.826—demonstrating strong sensitivity to subtle cheating behaviors. Trained for 288 epochs, it maintained real-time inference at 4.2 ms/image (~238 FPS) on an RTX 3090. In comparison, baseline YOLOv8 achieved 0.64 mAP@50, while YOLOv8+Mixed followed closely with 0.876. The LSTM+CNN model showed the highest accuracy (96.67%) but required a P100 GPU to reach 9 FPS, making it better suited for offline analysis. ViT, trained for just 15 epochs, offered good contextual performance (precision: 0.875, F1-score: 0.827) despite lower accuracy (81.42%) and slower speed. Efficient Channel Attention (ECA) proved effective in reducing false positives. These results highlight YOLOv8+ECA's real-time suitability and robustness for automated proctoring across practical exam scenarios.

Future Scope : Future enhancements of the offline AI proctoring system will focus on expanding multimodal capabilities, including audio input for whisper detection and keystroke monitoring to improve behavioral analysis. For low-resource environments, the YOLOv8+ECA model will be optimized with TensorRT for real-time performance on edge devices like the NVIDIA Jetson Nano. Incorporating Explainable AI (XAI) features such as attention heatmaps will enhance transparency and aid manual review. Multitarget tracking will support individualized behavior analysis by monitoring the type and frequency of suspicious actions. Few-shot and continual learning will enable adaptation to new cheating behaviors with minimal retraining. Integration with existing CCTV infrastructure can also enable real-time detection in institutional settings. Addressing dataset imbalances, tuning hyperparameters, and testing in varied conditions will improve accuracy and robustness. As not all anomalies indicate cheating, manual oversight remains essential, reinforcing the system's role as a support tool.

70102100018 SAMRIDHI RAJ SINHA
70102100051 ASMI PARIKH
70102100084 SAAKSHI JAIN

MENTOR: DR. SHUBHA PUTHRAN & DR. SUPRIYA AGARWAL
BRANCH : B.TECH CE

NMIMS
MUKESH PATEL SCHOOL OF TECHNOLOGY MANAGEMENT & ENGINEERING